



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE AGE RETURNS OF THE TWELFTH CENSUS
(1900).

By WILLIAM B. BAILEY, *Professor of Economics, Yale University*, and
JULIUS H. PARMELEE, *Special Agent, Bureau of the Census*.

Beginning with 1850, inquiry has been made at each census regarding the exact age in years of every person in the United States. The inquiry has usually referred to age at last birthday, although in 1890 the census schedule asked for "age at nearest birthday." In the Twelfth Census (1900) the inquiry was divided into two parts, the first asking for month and year of birth, the second for age at last birthday. The purpose of the addition of the question regarding date of birth was to increase the degree of accuracy of the statistics of age.

The suggestion that date of birth be ascertained appeared originally in the population report of the Eleventh Census (1890). The instructions for that census had especially cautioned enumerators against accepting round numbers in answer to the age query. "Particular attention should be paid to this," the instructions had read,* "otherwise it will be found when the results are aggregated in this office that a much more than normal number of persons have been reported as 20, 25, 30, 35, 40, etc., years of age, and a much less than normal at 19, 21, 24, 26, 29, 31, etc." As a result of this caution, concentration on multiples of five was somewhat lessened in 1890, but as it was not wholly done away with, the suggestion was made,† that to return ages by exact day, month, and year of birth would still further reduce such concentration.

The inquiry regarding the month and year of birth which was introduced into the schedule for the Twelfth Census conformed to this suggestion; and when analyses of the age re-

*Wright and Hunt, *History and Growth of the United States Census*, p. 188.

†Report on Population, Census of 1890, Part II, p. XV.

turns of that census indicated a gain in accuracy and a decrease in concentration on multiples of 5,* both results were at once ascribed to the inclusion of this additional inquiry. Dr. Allyn A. Young, who made a special study of age statistics for the Twelfth Census, said: "The addition of the inquiry as to date of birth has increased the general accuracy of the age returns."† He believed also, that the decrease in the amount of concentration on multiples of 5 was due to the addition of this same inquiry, and even that "as high a degree of accuracy would have been obtained if the inquiry as to 'age at last birthday' had been omitted."‡

In considering the advisability of retaining both inquiries regarding age in the population schedule of the Thirteenth Census, the question arose whether the inclusion in 1900 of the inquiry respecting date of birth did or did not result in improving the age statistics of that year. To assist in reaching a decision on this point, a critical study of the question was initiated by the Census Bureau, leading to the results and conclusions presented in this paper.

A careful inspection of ten bound volumes of the original census schedules of 1900 was made, to learn how large was the number of cases in which the reported year of age was inconsistent with the reported date of birth. It was found that out of 130,000 enumerations taken by census agents in five rural counties (one each in Kansas, Mississippi, New Hampshire, Oregon, and Pennsylvania) and five large cities (Boston, Chicago, New Orleans, New York and San Francisco), there were 12,526 cases of disagreement between date of birth and age recorded, or nearly one in ten.

In a certain number of these cases of disagreement the other data recorded on the schedules furnished conclusive evidence, or at least created a very strong presumption, as to which of the two age returns was the more reliable, one return being

*The estimated coefficient of error in the age returns fell from 7.5 per cent. in 1890 to 3.4 per cent. in 1900, or more than one half. The excess per cent. of concentration on multiples of 5 was reduced from 31.3 per cent. in 1890 to 19.8 per cent. in 1900. See Supplementary Analysis, Census of 1900, pp. 135, 137.

†Supplementary Analysis, Census of 1900, p. 141.

‡Supplementary Analysis of 1900, p. 137. See also Report on Population, Census of 1900, Part II, p. XXXV.

consistent with the other record data and the other inconsistent. Typical cases of this sort are the following, which are taken from Volume 145 of the bound schedules of 1900 for New York State and Volume 19 for Illinois:

Case 1

A—Son, born December, 1889, age 11 years.

B—Daughter, born May, 1890, age 10 years.

The ages reported by the enumerator were probably correct, and the year of birth of A incorrect. Otherwise A would be but five months older than his sister B.

Case 2

C—Born August, 1861, age 8 years.

As the schedule reported C to be attending school, the age reported by the enumerator was doubtless correct, and the year of birth incorrect.

Case 3

D—Born 1849, age 60.

E—Son of D, born 1859, age 40.

Here again the reported age of D was in all probability correct, and not the year of his birth; otherwise he would be but 10 years older than his son E.

Case 4

F—Wife, born 1872, age 38.

F in 1900 had been married 20 years. Her age, therefore, was correctly reported as 38, and the year of birth assigned to her was unquestionably wrong.

Case 5

G—Daughter, born 1882, age 13.

The reported age and year of birth of G are in disagreement. As G's parents, aged 35 and 33, respectively, had been married but 14 years, it is almost beyond question that her age is reported accurately, and the year of her birth inaccurately.

Case 6

H—Daughter, born October, 1899, age 10.

H was reported as attending school, which would hardly have been possible had the reported date of her birth been correct.

Case 7

J—Husband, born 1877, age 33.

J had been married 13 years and was the father of a boy of 11. His reported age is clearly more reliable than the reported year of his birth.

Case 8

K—Wife, born 1852, age 27.

Here the year of birth seems more trustworthy than the reported age in years. K was a foreign-born woman who had been in the country 28 years, and had been married 22 years of that time. Her reported age therefore, is evidently incorrect.

Of the cases selected as examples and cited above seven indicate an incorrect return of the year of birth and one an incorrect return of the year of age. In a count of 50 similar cases taken at random, or rather in the order in which they appeared in the bound volumes of the schedules consulted for this purpose, it was found that all but 6 indicated very conclusively that the return of age was correct, rather than that of year of birth. The fact that this proportion obtained in those cases where one could check the comparative accuracy of the two returns created a very strong presumption that in other cases, where there was no such check, the return of age was much more likely to be accurate than that of date of birth.

Upon a further inspection of the 12,526 cases of disagreement which had been checked off in the schedules, it was found that in more than two-thirds of these cases, the difference between the year of age return and the date of birth return was exactly one year, and that in every such case the person, according to the reported date of birth, was just one year

younger than his reported year of age. This typical relationship between the two returns in cases of disagreement could not have been accidental. It was evident there must be some causal connection — some explanation of the tendency to make the date of birth return one year younger than the year of age return. Further light was thrown upon the question when it was found that in all these cases almost without exception the birth month was one of the last seven months of the year. The following is a typical case: X is returned on the schedule as born in September, 1865, and as being 35 years of age. The census, it should be remembered, was taken as of June 1, 1900. Therefore if the date of birth was correct X was not 35 years of age on June 1, 1900, but only 34; if on the other hand he was, in fact, 35 years of age in June, 1900, and was born in September, then the year of his birth was not 1865, but 1864. The most probable explanation of these recurrent inconsistencies is that they result from a rather natural and obvious mistake made in computing the year of birth from the age given. X—or perhaps his wife or some other member of his family, X himself being away from home when the enumerator called — stated that he was 35 years old and was born in September. Then either he or his wife, or more likely the enumerator, subtracted 35 from 1900 and entered the year of birth as 1865. Of course it is possible that the year of birth may have been given directly as 1865, and the year of age then obtained by subtracting 1865 from 1900. This would result in the same inconsistency. To the writer, however, this seems unlikely, because it is his belief and observation that most people keep a better mental record of their age than of their year of birth, and will answer more readily and promptly an inquiry as to age in years than as to year of birth, whether the inquiry applies to themselves or to some relative, friend, or acquaintance.

This conclusion, however, as to the course followed in making the return of age in years does not rest wholly upon inherent probabilities. In the schedules examined, three cases were found where the enumerator had regularly entered the

year of birth with a different ink from that used in making the other entries, thus indicating very conclusively that he had computed and entered the year of birth at the close of his day's field work. Other enumerators may have done the same thing and yet left no trace of this procedure, because they used the same pen and ink that they had been using during the day, or because they copied each schedule entire before sending it in.

One can readily understand why the enumerator resorted to the practice of computing the year of birth if he found that people in general answered the question as to age in years readily enough, but on being asked the year of birth hesitated, became confused, tried to do a sum in subtraction, or went off to hunt up the family Bible. He naturally decided that the easiest way was to ask the age in years and month of birth—most persons know that—and then do his own subtracting in obtaining the year of birth. In the considerable proportion of cases he did it wrong, making the common mistake above described.

When these returns were received at Washington, the Census Bureau was forced to choose between the age return and the year of birth return in cases of disagreement. In view of the fact that the question as to date of birth had been inserted in the schedule on the assumption that it would bring out more accurate replies than the inquiry as to age, thereby reducing the concentration on quinquennial years, the office naturally instructed the clerks who were editing the schedules preparatory to tabulation to follow the year of birth, unless the accompanying data indicated that the year of birth was wrong—and in the rush of getting out the Census reports there was but little time for the slow work of checking by the accompanying data. At any rate, in all those cases where the difference was simply one year, the reported age was almost invariably changed in red ink so as to agree with the reported year of birth, thereby carrying over into the age tabulation the errors which the enumerator—or perhaps the person enumerated—had made in computing the year of birth.

It remains to consider what the effect was upon the age statistics of the Twelfth Census. A reduction of only one year in the ages of one-tenth of the population may not be regarded as a very serious matter, except perhaps as applied to children and young persons. Yet it may account for some things, and one thing which it seems to account for in part is the reduction in the concentration on the multiples of five in the age statistics of the Twelfth Census as compared with those of the Eleventh.

It is undoubtedly true that the enumeration of 1900 was more accurate than that of 1890, just as the enumeration of 1890 was more accurate than that of 1880. The gain in the accuracy of the age distribution of the population shown by the census of 1890 over that of 1880 was due almost entirely to greater care in the enumeration, and the numbers taken from the ages of concentration were distributed quite evenly over the years showing a deficit, as is shown in Table II on page 35. It is a question, however, whether the adjustments made in 1900 in the reported years of age, on the assumption that the reported year of birth was more reliable, did not assign an unduly large proportion of the numbers taken from the ages ending in 5 or 0 to the ages immediately preceding, namely, those ending in 4 or 9.

In order to determine this point the numbers of the total population at the different ages from 23 to 62 inclusive were taken for 1880, 1890, and 1900. A total was made of those at the ages of 23, 33, 43, and 53 at each census and totals also for ages ending in 4, 5, 6, 7, 8, 9, 0, 1 and 2. Percentages were then computed showing the proportion of total population at each age ending with a given integer.

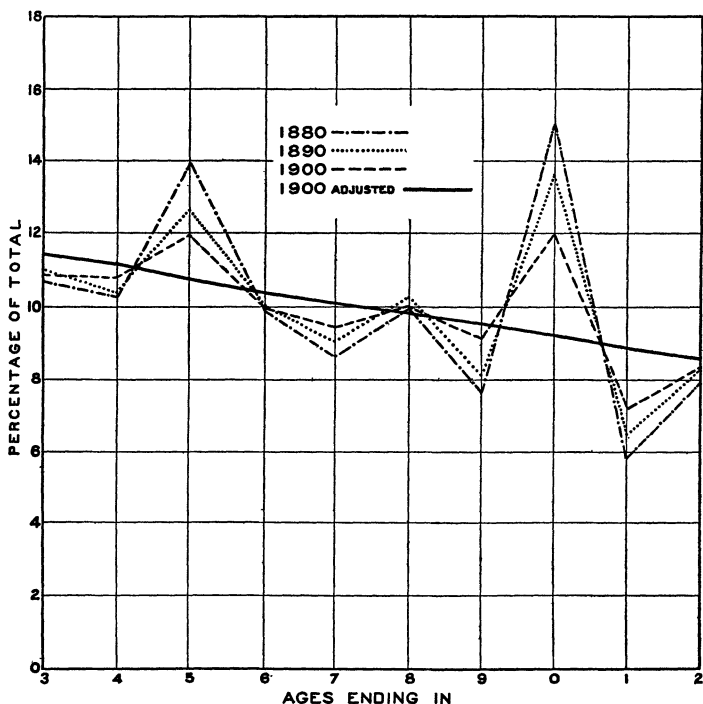
TABLE I.
DISTRIBUTION OF TOTAL POPULATION FROM 23 TO 62 YEARS OF AGE: 1880-1900.

Ages ending in	Adjusted ages: 1900*. Per cent.	1900.		1890.		1880.	
		Number.	Per cent.	Number.	Per cent.	Number.	Per cent.
Total.....	100.0	34,144,225	100.0	26,951,141	100.0	20,809,011	100.0
3.....	11.5	3,725,514	10.9	2,963,418	11.0	2,227,913	10.7
4.....	11.2	3,714,557	10.9	2,814,908	10.4	2,155,902	10.4
5.....	10.8	4,088,445	12.0	3,403,799	12.6	2,899,385	13.9
6.....	10.4	3,379,934	9.9	2,711,584	10.1	2,074,816	10.0
7.....	10.1	3,231,259	9.5	2,427,733	9.0	1,800,676	8.7
8.....	9.8	3,426,070	10.0	2,778,471	10.3	2,078,735	10.0
9.....	9.5	3,124,855	9.1	2,176,327	8.1	1,588,307	7.6
0.....	9.2	4,093,618	12.0	3,675,023	13.7	3,127,585	15.0
1.....	8.9	2,498,472	7.3	1,758,842	6.5	1,207,205	5.8
2.....	8.6	2,861,501	8.4	2,240,036	8.3	1,648,487	7.9

* Computed from Appendix A, Supplementary Analysis, Census of 1900, pp. 169 to 173. Here the errors that necessarily creep into census age returns, because of mis-statement of ages, are eliminated, or ironed out, by mathematical interpolation. See Allyn A. Young, "The Adjustment of Census Age Returns," Western Reserve University Bulletin, November, 1902.

From these data the following chart was drawn. This distribution of the adjusted or ideal population, it will be seen, approaches a straight line, while the line of each succeeding census from 1880 to 1900 more closely approximates this ideal. It may be noted in passing that in comparison with the adjusted age distribution, there was a concentration at each census on the ages ending in 5, 8, and 0. The concentration on ages ending in 8 was a very slight one, the years of marked concentration being those ending in 5 and 0.

AGE DISTRIBUTION OF THE POPULATION FROM 23 TO 62 YEARS
OF AGE: 1880-1900.



There was a reduction from 1880 to 1890 in the percentage of the population at ages ending in 5 from 13.9 per cent. to 12.6 per cent., or 1.3 per cent. At ages ending in 0 the reduction was from 15.0 per cent. to 13.7 per cent., or 1.3 per cent. There was thus a reduction in the ages of concentration amounting to 1.3 plus 1.3, or 2.6 per cent. There was a similar reduction of 2.3 per cent. between 1890 and 1900.

These reductions are, of course, offset by increases of 2.6 per cent. and 2.3 per cent., respectively, distributed over the remaining ages. The following table shows the distribution of the numbers taken from the ages of concentration.

TABLE II.

Ages ending in	Increase or decrease of per cent.	
	1890 to 1900.	1880 to 1890.
0 and 5.....	-2.3	-2.6
1 and 6.....	0.6	0.8
2 and 7.....	0.6	0.7
3 and 8.....	-0.4	0.6
4 and 9.....	1.5	0.5

The distribution of the numbers taken from the ages of concentration was much more regular in the census of 1890 than in that of 1900. Nearly three-fourths of the amount taken from the ages ending in 5 and 0 in 1900 was credited to ages ending in 4 and 9. While the concentration on ages ending in 5 and 0 was thus considerably reduced, the reduction was accompanied by a biased distribution of the numbers transferred to other ages. It was not to be expected, of course, that this distribution would be absolutely regular. As the general tendency of people is, to understate ages,* any concentration on years ending in 0 and 5 is drawn largely from the years immediately succeeding, i. e., those ending in 1 and 6; and any increase in the accuracy of age returns will,

*See Supplementary Analysis, Census of 1900, p. 130.

in lessening the concentration on years ending in 0 and 5, add more to the years ending in 1 and 6 than other years. This was true in 1890, as Table II shows. In 1900, also, the increase for the years ending in 1 and 6 was a comparatively large one, although not half so large as the increase for the years ending in 4 and 9. The latter increase was undoubtedly brought about by means of the clerical adjustments considered above, which consisted largely in reducing the reported age one year, to conform to the reported date of birth.

The problem was approached from still another angle. It has already been stated that a careful examination of 130,000 representative urban and rural census enumerations of 1900 disclosed lack of agreement between reported age and date of birth in nearly one-tenth of the age returns. There were 12,526 such cases of disagreement, which it was necessary to adjust in some manner. Of the 12,526 clerical adjustments made 8,851, or 70.7 per cent., reduced the reported ages by one year and 950,† or 7.6 per cent., by one month, while the remainder comprised all manner of emendations.

The 8,851 corrections representing reductions of one year were found to be of the following ten types.

TABLE III.

Reductions		Number.
From ages ending in	To ages ending in	
0	9	1,150
1	0	588
2	1	969
3	2	851
4	3	866
5	4	1,102
6	5	893
7	6	856
8	7	882
9	8	694

†These cases were children less than one year old, whose ages were reported in terms of months.

It will be seen that the ages ending in 0 lost 1,150 to the ages ending in 9, and received 588 from the ages ending in 1; the net loss on the part of the ages ending in 0 was therefore 562. Similarly, the ages ending in 5 suffered a net loss of 209. For ages ending in 0 and 5 the net loss, then, was 562 plus 209, or 771. The net gain or loss for all the age groups is shown in the following table:

TABLE IV.

Ages ending in	Net gain or loss
0 and 5	-771
1 and 6	+344
2 and 7	- 92
3 and 8	-173
4 and 9	+692

Thus the net loss from the ages ending in 0 and 5 is nearly balanced by the net gain to the ages ending in 4 and 9; in other words, the one-year adjustments did, it is true, reduce the concentration on multiples of 5, but loaded that reduction almost entirely upon the ages ending in 4 and 9, where only part of it properly belonged. The net increase in the ages ending in 1 and 6 was practically balanced by the reductions in the remaining ages.

The excess per cent. of concentration on multiples of 5 was estimated by Dr. Young as follows:*

Census of 1880	44.8 per cent.
Census of 1890	31.3 per cent.
Census of 1900	19.8 per cent.

If the relative number of one-year reductions in the 130,000 enumerations examined was typical of all the census returns for 1900,—and the schedules examined were representative of urban and rural conditions in all sections of the country,—the reductions made for the whole United States must have

*Supplementary Analysis, Census of 1900, p. 137.

been responsible for a decrease of 2.9 per cent.† in the concentration on multiples of 5. If the reductions had not been made, therefore, the excess concentration in 1900 would have been 22.7 per cent., and not 19.8 per cent.

Looking at the question from every point of view, then, it seems clear that in the ages as actually returned on the Twelfth Census schedules the decrease of concentration was not so great as it appeared to be after these returns had been adjusted to agree with the return of the year of birth; and since the return of the year of birth was more inaccurate than the return of the age in years, it follows that the gain in accuracy of the age statistics of the Twelfth Census was somewhat smaller than has been supposed, or was, in other words, to some extent a delusion.

As the actual distribution of the population by age groups comes in succeeding census enumerations to correspond more and more nearly to that of an adjusted or ideal population, the gain in accuracy will become gradually smaller. The concentration on certain ages will probably never disappear, as it is a function of two variables: the care of the enumerator in asking the questions and recording the answers, and the ignorance or insincerity of the person questioned. It is desirable that each census should show improvement over its predecessors, but it seems probable that the improvement in the census of 1900 was not so great as has been believed. From the facts adduced in the preceding pages the following conclusions were drawn:

(1) That the inclusion in the census schedule in 1900 of an inquiry as to date of birth was rendered valueless in a considerable proportion of the cases by the practice of enumerators, or others, of computing date of birth from the age given.

(2) That in a considerable proportion—perhaps one tenth—of the cases the year of birth was incorrectly computed, the

†This figure was obtained as follows: The number of ages ending in 0 and 5 is normally one fifth of all the ages in any given age series. The normal number of such ages in the 130,000 cases under consideration would thus be 26,000. The ratio of 771, the net loss on the part of such ages, as shown in Table IV, to 26,000, is 2.9 per cent.

extent of the inaccuracy being in a large majority of instances exactly one year.

(3) That the assumption on the basis of which the schedules were corrected in the office of the census, i. e., that where the date of birth and the age in years were in disagreement, the former was the more reliable, is not borne out by a careful study of the internal evidence in the schedules themselves.

(4) That the adjustments made on this assumption resulted in an extensive reduction of ages by one year, which by transferring a large number of the ages ending in 5 and 0 to the ages ending in 4 and 9, fostered the belief that the improvement in the age returns of 1900 was somewhat greater than was probably the actual case.

In view of the conclusions just presented, and especially in view of the first conclusion, that the inquiry as to date of birth played little or no part, either in increasing the accuracy of the age returns of the Twelfth Census, or in reducing the concentration on years ending in the integers 5 and 0, it was decided to eliminate the query regarding date of birth from the population schedule of the Thirteenth Census, and to retain only the question regarding age at last birthday.